FULL-INFORMATION ITEM FACTOR ANALYSIS REVISION(U)
NATIONAL OPINION RESEARCH CENTER CHICAGO IL
R D BOCK ET AL. JUL 86 MRC-85-1-REV N00014-83-C-0203
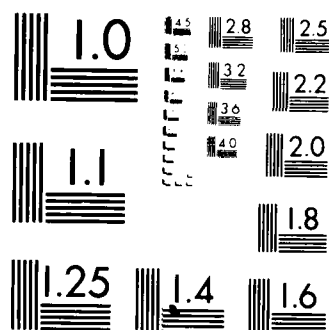
UNCLASSIFIED
F/G 12/1
NL

		1.0			2.8		2.5
					3.2		2.2
					3.6		
		1.1			4.0		2.0
							1.8
		1.25		1.4		1.6

MICROCOPY RESOLUTION TEST CHART
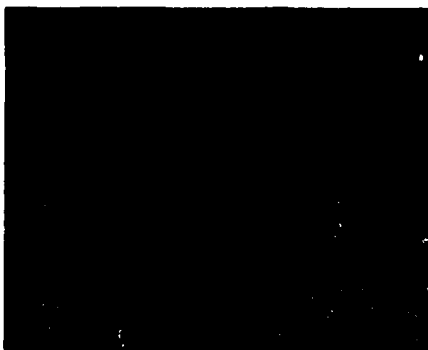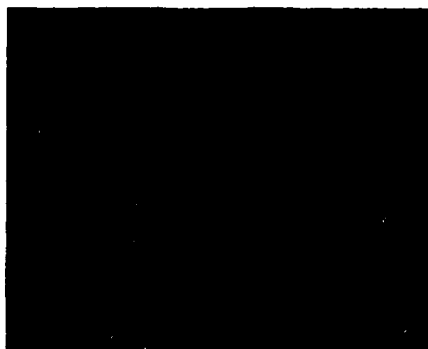
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A170 602

NFORMATION ITEM FACTOR ANALYSIS

R. Darrell Bock
University of Chicago

Robert Gibbons
University of Illinois

and

Eiji Muraki
National Opinion Research Center
MRC Report #85-1 (Revised)

DTIC FILE COPY

NORC

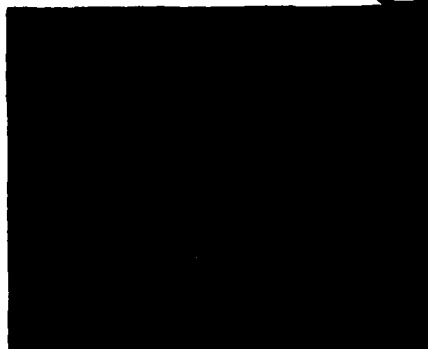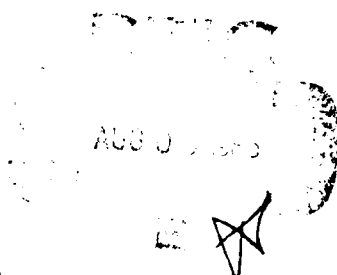UNIVERSITY OF CHICAGO   86 8 7 105

12

FULL-INFORMATION ITEM FACTOR ANALYSIS

R. Darrell Bock
University of Chicago

Robert Gibbons
University of Illinois

and

Eiji Muraki
National Opinion Research Center
MRC Report #85-1 (Revised)

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION		1b RESTRICTIVE MARKINGS
Unclassified		NONE

2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
Unclassified	Approved for public release
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	

4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
MRC Report #85-1(Revised)	

6a NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Methodology Research Ctr./NORC		Office of Naval Research (Code 442PT)

6c ADDRESS (City, State, and ZIP Code)	7b ADDRESS (City, State, and ZIP Code)
6030 S. Ellis Avenue Chicago, IL 60637	800 North Quincy Street Arlington, VA 22217-5000

8a NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Office of Naval Research		N00014-83-C-0283 N00014-83-C-0457

8c. ADDRESS (City, State, and ZIP Code)

800 North Quincy Street
Arlington, VA 22217-5000

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
62763N	RF63521	RF63521803	NR 475-018

11 TITLE (Include Security Classification)

Full-Information Item Factor Analysis

12 PERSONAL AUTHOR(S)
R. Darrell Bock, Robert Gibbons, and Eiji Muraki

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Revised Final	FROM _____ TO _____	July, 1986	

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A method of item factor analysis based on Thurstone's multiple factor model and implemented by marginal maximum likelihood estimation and the EM algorithm is described. Statistical significance of successive factors added to the model is treated by the likelihood ratio criterion. Provisions for effects of guessing on multiple choice items, and for omitted and not reached items, are included. Bayes constraints on the factor loadings are found to be necessary to suppress Heywood cases. Numerous applications to simulated and real data are presented to substantiate the accuracy and practical utility of the method. Analysis of the power tests of the Armed Services Vocational Battery shows statistically significant departures from unidimensionality in five of eight tests.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
R. Darrell Bock	(312) 962-9859	

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete

Abstract

A method of item factor analysis based on Thurstone's multiple factor model and implemented by marginal maximum likelihood estimation and the EM algorithm is described. Statistical significance of successive factors added to the model is tested by the likelihood ratio criterion. Provisions for effects of guessing on multiple choice items, and for omitted and not reached items, are included. Bayes constraints on the factor loadings are found to be necessary to suppress Heywood cases. Numerous applications to simulated and real data are presented to substantiate the accuracy and practical utility of the method. Analysis of the power tests of the Armed Services Vocational Battery shows statistically significant departures from unidimensionality in five of the eight tests.
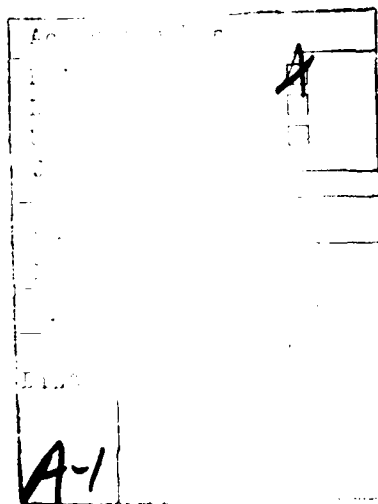
FULL-INFORMATION ITEM FACTOR ANALYSIS

R. Darrell Bock
University of Chicago

Robert Gibbons
University of Illinois

and
Eiji Muraki
National Opinion Research Center

July 16, 1986

Strictly speaking, any test reported in a single score should consist of items drawn from a one-dimensional universe. Only then is it a matter of indifference which items are presented to the examinee. This interchangeability of items is especially important in adaptive testing, where different examinees are presented different items.

Of the various methods that have been proposed for investigating the dimensionality of item sets, the most sensitive and informative is item factor analysis. It alone is capable of analyzing relatively large numbers of items jointly and symmetrically, and of assigning items to particular dimensions when multiple factors are found. It can also reveal patterns of common item content and format that may have interesting cognitive interpretations.

Regrettably, the methods of item factor analysis heretofore available have not been entirely satisfactory technically. Although conventional multiple factor analysis of the matrix of phi coefficients is straightforward computationally, it is well known to introduce spurious factors when the item difficulties are not uniform. This problem is alleviated by using tetrachoric correlations in place of phi coefficients, but this strategy also encounters difficulties. The matrix of sample tetrachoric correlation coefficients is almost never positive definite, so the common factor model does not strictly apply. Although present methods of calculating the tetrachoric coefficents are fast and generally accurate (Divgi, 1979), they become unstable as the values approach $+1$ or -1. When an observed frequency in the four-fold table for a pair of items is zero, the absolute value of an element in the item correlation matrix becomes 1, thus producing a Heywood case. These problems are exacerbated when the coefficients are corrected for guessing (Carroll, 1945).

The limitations of the item factor analysis based on tetrachoric correlation coefficients have been overcome to a considerable extent by the generalized least squares

1

(GLS) method (Cristoffersson, 1975; Muthén, 1978). Because this method allows for the large sample variance of the estimated coefficients, instabilities at the extremes are less of a problem. The GLS method requires, however, the generating and inverting of the asymptotic covariance matrix of the estimated tetrachoric coefficients; it thus becomes extremely heavy computationally as the number of items increases. At present, its practical upper limit is about 20 items (Muthén, 1984).

It is of some interest, therefore, that Bock and Aitkin (1981) have introduced a method of item factor analysis, based directly on item response theory, that does not require calculation of inter-item correlation coefficients and is not strongly limited by the number of items. Although the computations in their method increase exponentially with the number of factors, they increase only linearly with the number of items. The practical limit of the number of factors is five, which is sufficient for most item analysis applications, while 60 to 100 items is not excessive.

Because the Bock-Aitkin approach uses as data the frequencies of all distinct item response vectors, it is called "full-information" item factor analysis (Bartholomew, 1980). It contrasts with the limited information methods of Cristoffersson and Muthén based on low-order joint occurrence frequencies of the item scores. The purpose of the present paper is to present in more detail the derivation of the full-information factor analysis, discuss technical problems of its implementation, and describe our experience with the procedure in a number of simulated and real data sets.

1   DERIVATION AND STATISTICAL METHODS

Bock & Aitkin (1981) apply Thurstone's multiple factor model to item response data by assuming that the m-factor model,

$$Y_{ij} = \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + c\ldots + \alpha_{jm}\theta_{mi} + v_{ij}, \tag{1}$$

describes not a manifest variable, but an unobservable "response process." The process generates a correct response of person i to item j when y_{ij} equals or exceeds a threshold, γ_j, and yields an incorrect response, otherwise. Thus, on the assumption that v_{ij} is an unobservable random variable distributed $N(0, \sigma_j^2)$, the probability of an item score, $x_{ij} = 1$, indicating a correct response to item j from person i, with abilities $\theta_i = [\theta_{1i}, \theta_{2i}, \ldots, \theta_{mi}]$, is

$$
\begin{aligned}
P(x_{ij} = 1 \mid \theta_i) &= \frac{1}{\sqrt{(2\pi)}\sigma} \int_{\gamma_j}^{\infty} \exp\left[-1/2\left(\frac{Y_{ij} - \sum_k^m \alpha_{jk}\theta_{ki}}{\sigma_j}\right)\right] dy_{ij} \\
&= \Phi[(\gamma_j - \sum_k^m \alpha_{jk}\theta_{ki})/\sigma_j]
\end{aligned}
$$

2

$$= \Phi_j(\theta_i) \qquad (2)$$

The conditional probability of the item score $x_i = 0$, indicating an incorrect response, is the complement, $1 - \Phi_j(\theta)$. In other words, the conditional response probability is given by a normal ogive model. Note that (1) is a "compensatory" model: greater ability in one dimension makes up for lesser ability in some other dimension. Nothing, however, prevents the methods discussed here from being applied to an "interactive" model such as

$$y_{ij} = \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \alpha_{j12}\theta_{1i}\theta_{2i} + \cdots + \alpha_{jmp}\theta_{mi}\theta_{pi} + v_{ij} \qquad (3)$$

1.1   Estimation of the Item Thresholds and Factor Loadings

Like maximum likelihood factor analysis for measured variables (Jöreskog, 1967), the Bock-Aitkin method of estimating parameters of an item-response model assumes that the data have been obtained from a sample of persons drawn from a population with some multivariate distribution of ability. Provisionally, we will assume that the distribution is $\theta \sim N(0, \mathbf{I})$, but this assumption can be relaxed to allow for correlated factors and non-normal distributions. We also adopt the convention of factor analysis that y_j is distributed with mean zero and variance one, so that

$$\sigma_j^2 = 1 - \sum_{k=1}^{m} \alpha_{jk}^2 \qquad (4)$$

On these assumptions, the marginal probability of the binary response pattern is given by the multiple integral,

$$
\begin{aligned}
\tilde{P}_\ell &= P(\mathbf{x} = \mathbf{x}_\ell) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} [\Phi_j(\theta)]^{x_{\ell j}} [1 - \Phi_j(\theta)]^{1-x_{\ell j}} g(\theta) d\theta \\
&= \int_\theta L_\ell(\theta) g(\theta) d\theta \qquad (5)
\end{aligned}
$$

Numerical approximations of these integrals may be obtained by the m-fold Gauss-Hermite quadrature,

$$\tilde{P}_\ell = \sum_{q_m=1}^{Q} \cdots \sum_{q_m=2}^{Q} \sum_{q_1=1}^{Q} L_\ell(\mathbf{X}_k) A(X_{q_1}) A(X_{q_2}) \cdots A(X_{q_m}), \qquad (6)$$

where $\mathbf{X}_k$ is a quadrature point in m dimensional space and the corresponding weight is the product of weights for quadrature in the separate dimensions as shown.

3

Equation (6) applies, of course, only to uncorrelated factors. It is an example of the so-called "product formula" for numerical integration and has the disadvantage that the number of terms in the sum is an exponential function of the number of dimensions. Fortunately, the number of points in each dimension can be reduced as the dimensionality is increased without impairing the accuracy of the approximations. Thus, factor analysis with five factors can be performed with good accuracy with as few as three points per dimension. In that case, $3^5 = 243$ quadrature points are required, and the solution is accessible to a fast computer.

Given the frequencies, r_ℓ of the response patterns, x_ℓ for n items and a sample of N persons, the number of distinct pattern is $s \leq min(2^n, N)$, and the probability of the sample is

$$L_M = P(X) = \frac{N!}{r_1! r_2! \ldots, r_s!} \bar{P}_1^{r_1} \bar{P}_2^{r_2} \ldots, \bar{P}_s^{r_s} \tag{7}$$

Then the maximum likelihood estimates of the threshold and factor loadings are those values that maximize (7). To simplify the expression of the likelihood equations, it is convenient to express the model in terms of the intercept and slopes of the response function and to write

$$\frac{\gamma_j - \sum_k^m \alpha_{jk} \theta_{ki}}{\sigma_j} = c_j - \sum_k^m a_{jk} \theta_{ki} \tag{8}$$

From MML estimates of c_j and a_{jk} the latter, MML estimates of γ_j and α_{jk} may be obtained by

$$\hat{\gamma}_j = -\hat{c}_j / \hat{d}_j \tag{9}$$

and

$$\hat{\alpha}_{jk} = \hat{a}_{jk} / \hat{d}_j , \tag{10}$$

where

$$\hat{d}_j = (1 + \sum_k^m \hat{a}_{jk}^2)^{1/2} \tag{11}$$

Notice that the item threshold in this model is not an invariant statistic: it depends upon the distribution of ability in the sample and is on the response process dimension rather than on an ability dimension. The invariant location parameter of the one dimensional model does not exist in the multidimensional case: the value of one ability that corresponds to a given probability of correct response is a linear function of the other abilities.

The likelihood equation for a general item parameter, ν_j, is:

$$\frac{\partial \log L_M}{\partial \nu_j} = \sum_{\ell=1}^{s} \frac{r_\ell}{\tilde{P}_\ell} \cdot \frac{\partial \tilde{P}_\ell}{\partial \nu_j}$$

$$= \sum_{\ell}^{s} \frac{r_\ell}{\tilde{P}_\ell} \int_\theta \frac{L_\ell(\theta)}{[\Phi_j(\theta)]^{x_{\ell j}}[1-\Phi_j(\theta)]^{1-x_{\ell j}}} \cdot \frac{\partial\{[\Phi_j(\theta)]^{x_{\ell j}}[1-\Phi_j(\theta)]^{1-x_{\ell j}}\}}{\partial \nu_j} g(\theta)d\theta$$

$$= \sum_{\ell}^{s} \frac{r_\ell}{\tilde{P}_\ell} \int_\theta \left(\frac{x_{\ell j} - \Phi_j(\theta)}{\Phi_j(\theta)[1-\Phi_j(\theta)]}\right) L_\ell(\theta) \frac{\partial \Phi_j(\theta)}{\partial \nu_j} g(\theta)d\theta$$

$$= \int_\theta \frac{\bar{r}_j - \bar{N}\Phi_j(\theta)}{\Phi_j(\theta)[1-\Phi_j(\theta)]} \cdot \frac{\partial \Phi_j(\theta)}{\partial \nu_j} g(\theta)d\theta, \tag{12}$$

where

$$\bar{r}_j = \sum_{\ell=1}^{s} \frac{r_\ell x_{\ell j} L_\ell(\theta)}{\tilde{P}_\ell} \tag{13}$$

and

$$\bar{N} = \sum_{\ell=1}^{s} \frac{r_\ell L_\ell(\theta)}{\tilde{P}_\ell} \tag{14}$$

The multiple integral in this equation may be evaluated numerically by repeated Gauss-Hermite quadrature as follows:

$$\sum_{q_m}^{Q} \cdots \sum_{q_2}^{Q} \sum_{q_1}^{Q} \frac{\bar{r}_{j,q_1 q_2 \cdots q_m} - \bar{N}_{q_1 q_2 \cdots q_m} \Phi_j(\mathbf{X})}{\Phi_j(\mathbf{X})[1-\Phi_j(\mathbf{X})]} \cdot \frac{\partial \Phi_j(\mathbf{X})}{\partial \nu_j} A(X_{q_1}) A(X_{q_2}) \cdots A(X_{q_m})$$

The expected frequency $\bar{r}_{j,q_1 q_2 \cdots q_m}$ is an entry in a Q^m dimensional array in which each cell corresponds to an m-tuple of quadrature points for a given item. The entries in this table are the numbers of examinees with abilities equal to the vector $\mathbf{X}_\ell$ who are expected to respond correctly to the item, given the sample data.

The quantity $\bar{N}_{q_1 q_2 \cdots q_m}$ is the margin of this array summed over items; it is the expected number of persons with ability $\mathbf{X}_\ell$ and is normalized to the sample size.

These equations correspond to the steps in the so-called "EM" algorithm for marginal maximum likelihood estimation as given by Dempster, Laird, and Rubin (1977). Equations (13) and (14) comprise the E-step, in which expectations of "complete data" statistics are computed conditional on the "incomplete data." Equation (12) is the M-step, in which conventional maximum likelihood estimation is carried out using the expectations in place of complete data statistics. Because the expectations depend upon the parameters to be estimated, however, the calculations must be carried out iteratively. Given starting values for the parameters,

5

a Q^m table of expected frequencies, $\bar{r}_{j,q_1 q_2 \cdots q_m}$, giving the numbers of correct responses at each point, X_k, is built up for each item by distributing corresponding item score weighted by the posterior probability of the response pattern, x_ℓ, at point x_k. Similarly, $\bar{N}_{q_1 q_2 \cdots q_m}$ is obtained as the sum of the weights at each point. From these statistics, improved estimates of the item parameters are obtained in the M-step by applying the appropriate maximum likelihood solution to the table corresponding to the item in question. In the present case, any standard procedure for multiple probit analysis will suffice for the M-step. But the procedure is general for any item-response model; if a logistic response model were assumed, a multiple logit analysis would appear in the M-step. A version of this procedure based on the normal response model is implemented in the TESTFACT program of Wilson, Wood and Gibbons (1984).

1.2 Testing the Number of Factors

If the sample size is sufficiently large that all 2^n possible response patterns have expected values greater than one or two, the chi-square approximation for the likelihood ratio test of fit of the model relative to the general multinomial alternative is

$$G^2 = 2 \Sigma r_\ell \ln(r_\ell / N \tilde{P}_\ell), \tag{15}$$

where $\tilde{P}_\ell$ is computed from the maximum likelihood estimates of the item parameters. The degrees of freedom are

$$2^n - n(m+1) + m(m-1)/2$$

In this case, the goodness of fit test can be carried out after performing repeated full-information analyses, adding one factor at a time. When G^2 falls to insignificance, no further factors are required.

When the number of patterns is larger than the sample size, however, some of the expected frequencies may be near zero. In this case, (15), or other approximations to the likelihood ratio statistic for goodness-of-fit, becomes inaccurate and cannot be relied on. Haberman (1977) has shown, however, that the difference in these statistics for alternative models is distributed in large samples as chi-square, with degrees of freedom equal to the difference of respective degrees of freedom, even when the frequency table is sparse. Thus, the contribution of the last factor added to the model is significant if the corresponding change of chi-square is statistically significant. We investigate properties of the change chi-square statistic empirically in sections 3 and 4.

6

2 IMPLEMENTATION OF THE FULL-INFORMATION FACTOR ANALYSIS

Typically, EM solutions converge so slowly that devices such as Ramsay's (1975) acceleration method must be used to speed up the computation. For the same reason, it is important that the solution begin from accurate starting values. A good strategy for obtaining starting values is to perform a principal factor analysis, with communality iteration, on the matrix of tetrachoric correlations of the items in question. The tetrachoric correlation matrix should be corrected for guessing, and for missing values, and conditioned to be positive-definite.

Since the factors of the principal factor analysis are orthogonal, their loadings are suitable for the full-information solution after conversion to item intercepts and slopes. Item intercept and slope estimates based on the full-information method are then converted again into factor loadings. The resulting full-information factor pattern can be rotated orthogonally to the varimax criterion (Kaiser, 1958) and, with the varimax solution as target, rotated obliquely by the promax method (Hendrickson and White, 1964). The promax pattern is especially useful for identifying one-dimensional subsets of items into which a multidimensional set may be partitioned in order to measure abilities in the separate dimension.

2.1 Correction For Guessing

Carroll (1945, 1983) has warned against artifacts introduced into item factor analysis by guessing on multiple choice items. To suppress these effects, he proposes corrections to the four-fold tables from which the tetrachoric correlations are computed. In the full-information analysis, a similar solution results from substituting, for the normal ogive response function, the guessing model with lower asymptote g_j:

$$\Phi_j^*(\theta) = g_j + (1 - g_j)\Phi_j(\theta), \tag{16}$$

The lower asymptotes for the items may be estimated by marginal maximum likelihood along with the intercept and slope parameters, possibly with a prior distribution assumed for g_j in the M-step.

If the item response model with guessing parameter is used for the full-information factor analysis, the tetrachoric correlation matrix used to produce starting values of the parameters must be corrected for guessing prior to the principal factor analysis. To express Carroll's correction method in terms of the proportions in the 2 x 2 table, let us denote by g_i and g_j the probability of chance success on items i and j, respectively. Denote by π_{ij} the observed proportions in the original 2 x 2 table, and

by π'_{ij} the proportions in the corrected 2 x 2 table. Thus, the original and corrected contingency tables may be expressed as in Table 2-1 and 2-2, respectively.

TABLE 2-1

ORIGINAL PROPORTIONS OF SUBJECTS
PASSING AND FAILING ITEMS i AND j

		Item j		
		Pass	Fail	Total
	Pass	π_{11}	π_{10}	$\pi_{1\cdot}$
Item i	Fail	π_{01}	π_{00}	$\pi_{0\cdot}$
	Total	$\pi_{\cdot 0}$	$\pi_{\cdot 0}$	1

TABLE 2-2

CORRECTED PROPORTIONS OF SUBJECTS
PASSING AND FAILING ITEMS i AND j

		Item j		
		Pass	Fail	Total
	Pass	π'_{11}	π'_{10}	$\pi'_{1\cdot}$
Item i	Fail	π'_{01}	π'_{00}	$\pi'_{0\cdot}$
	Total	$\pi'_{\cdot 1}$	$\pi'_{\cdot 0}$	1

The guessing parameter is the probability of observing a correct response when, given the true state of mastery for the item, the response should be failure. Thus, the observed proportion of passing is the sum of the proportion of the true state of mastery and the joint proportions of the corresponding guessing and the true failure state. Therefore, we obtain

$$\pi_{1\cdot} = \pi'_{1\cdot} + g_i \pi'_{0\cdot}$$
$$\pi_{\cdot 1} = \pi'_{\cdot 1} + g_j \pi'_{\cdot 0}$$
$$\pi_{11} = \pi'_{11} + g_i \pi'_{01} + g_j \pi'_{10} + g_i g_j \pi'_{00}$$

and

$$\pi'_{11} + \pi'_{01} + \pi'_{10} + \pi'_{00} = 1 \tag{17}$$

From Equations (17), we solve the corrected proportions π' in terms of the observed proportion π and guessing parameters g as follows:

$$\pi'_{00} = \pi_{00}/w_i w_j$$
$$\pi'_{01} = (w_j \pi_{01} - g_j \pi_{00})/w_i w_j$$
$$\pi'_{10} = (w_j \pi_{10} - g_i \pi_{00})/w_i w_j$$

and

$$\pi'_{11} = 1 - \pi_{00} - \pi_{01} - \pi_{10} \qquad (18)$$

where $w_i = 1 - g_i$ and $w_j = 1 - g_j$.

To correct the item statistics for chance success, we proceed as follows. The conversion of the kth factor loading, α_{jk}, to the provisional slope estimate, a_{jk}, is

$$a_{jk} = \alpha_{jk}/\sigma_j, \qquad (19)$$

where

$$\sigma_j^2 = 1 - \Sigma \, \alpha_{jk}^2$$

The provisional intercept estimate, c_j, is computed from σ_j and the standard difficulty, δ_j, by

$$c_j = \delta_j/\sigma_j, \qquad (20)$$

since

$$\sigma_j = d_j^{-1}$$

The standard difficulty, δ_j, is the inverse normal transform of the item facility, π_j, which is measured by the proportion of individuals passing item j. The corrected facility, π'_j, is computed from

$$\pi'_j = 1 - (1 - \pi_j)/(1 - g_j). \qquad (21)$$

2.2 Correction For Omitted Responses

A disadvantage with Carroll's correction is that it fairly often produces a zero or negative value in an off-diagonal element of the four-fold table. If all omitted responses are recoded as incorrect responses, the observed proportions, π_{10}, π_{01}, and π_{00}, tend to be inflated. Since the positive corrected proportions are obtained only if $\pi_{00}/\pi_{0.} \leq w_j$ and $\pi_{00}/\pi_{.0} \leq w_i$, negative corrected proportions are the likely result. This problem is almost always encountered because omitted responses are frequently found in cognitive testing. A possible solution for this problem is to allocate omitted responses to the categories of correct and incorrect responses as

9

TABLE 2-3

OBSERVED FREQUENCIES OF SUBJECTS PASSING, FAILING, AND OMITTING ITEMS i AND j

		Item j			
		Pass	Fail	Omit	Total
	Pass	n_{11}	n_{10}	n_{1z}	$n_{1\cdot}$
	Fail	n_{01}	n_{00}	n_{0z}	$n_{0\cdot}$
Item i					
	Omit	n_{z1}	n_{z0}	n_{zz}	$n_{z\cdot}$
	Total	$n_{\cdot 1}$	$n_{\cdot 0}$	$n_{\cdot z}$	$n_{\cdot\cdot}$

shown below. This correction for omits must be made before the correction for guessing.

Let the observed frequencies in the 3 x 3 array whose categories are pass, fail, and omit, be expressed as in Table 2-3.

If the proportions of correct and incorrect responses based on non-omitted responses are denoted by p's and q's respectively, they are computed by

$$p_i = (n_{11} + n_{10})/N_{\cdot\cdot}$$
$$q_i = (n_{01} + n_{00})/N_{\cdot\cdot}$$
$$p_j = (n_{11} + n_{01})/N_{\cdot\cdot}$$

and

$$q_j = (n_{10} + n_{00})/N_{\cdot\cdot} \tag{22}$$

where

$$N_{\cdot\cdot} = n_{11} + n_{10} + n_{01} + n_{00}$$

If we can assume that omitted responses can be reallocated to correct and incorrect responses proportional to p and q, the following corrected frequencies n'_{ij} are obtained:

$$n'_{11} = n_{11} + p_j n_{1z} + p_i n_{z1} + p_i p_j n_{zz}$$
$$n'_{10} = n_{10} + q_j n_{1z} + p_i n_{z0} + p_i q_j n_{zz}$$
$$n'_{01} = n_{01} + p_j n_{0z} + q_i n_{z1} + q_i p_j n_{zz}$$

10

and

$$n'_{00} = n_{00} + q_j n_{0x} + q_i n_{x0} + q_i q_j n_{xx} \tag{23}$$

Marginal frequencies are computed by

$$
\begin{aligned}
n'_{1\cdot} &= n_{1\cdot} + p_i n_{x\cdot} \\
n'_{0\cdot} &= n_{0\cdot} + q_i n_{x\cdot} \\
n'_{\cdot 1} &= n_{\cdot 1} + p_j n_{\cdot x}
\end{aligned}
$$

and

$$n'_{\cdot 0} = n_{\cdot 0} + q_j n_{\cdot x} \tag{24}$$

Therefore,

$$n'_{1\cdot} + n'_{0\cdot} = n'_{\cdot 1} + n'_{\cdot 0} = n_{\cdot\cdot},$$

because

$$p_i + q_i = p_j + q_j = 1.$$

2.3   Preliminary Smoothing of the Tetrachoric Correlation Matrix

Although the correction for omits makes the calculation of most of the tetrachoric correlations possible, there are still occasional instances in large matrices where a value close to 0 appears in the minor diagonal of the 2 x 2 table. Since no admissible coefficient can be computed from such a table, some method of imputing a value is required. A reasonable approach is to assume that the matrix of tetrachoric correlations is dominated by a single factor. In that case, Thurstone's centroid formula applied to the valid correlations can be used to estimate the item factor loadings from which the missing coefficients can be calculated. Because the full-information analysis uses the tetrachoric correlations only for starting values, no bias of the solution results from these imputations.

To be analyzed by the MINRES method (Harman, 1976), the tetrachoric matrix must be positive-definite. The corrected matrix obtained through the centroid method, on the other hand, may have zero and negative roots. Therefore, a preliminary "smoothing" of the tetrachoric correlation coefficient matrix is needed before the principal factor analysis is carried out. The smoothed tetrachoric correlation matrix is produced from the eigenvectors associated with the positive roots, after renorming the sum of the roots to equal the number of items. The reproduced positive definite tetrachoric correlation matrix is then analyzed by the MINRES method to obtain good starting values for the full-information factor analysis.

11

2.4 Constraints On Item Parameter Estimates

An undesirable feature of maximum likelihood factor analysis is its tendency to produce Heywood cases, i.e., boundary solutions in which the uniqueness is zero for one or more variables. These cases also occur in full-information item factor analysis, the symptom being one or more continually increasing item slopes as the EM cycles continue.

One way of handling this problem is to assume a restricted prior distribution on some of the item parameters and to employ a maximum posteriori (MAP) estimation, i.e., to maximize the posterior probability density of the parameters rather than the likelihood. Martin and McDonald (1973) assume an exponential distribution for the uniqueness, while Lee (1981) employs an inverted gamma prior. Mislevy (1984) suggests that, since the uniqueness is bounded between 0 and 1, the beta prior

$$f(\sigma_j^2) = B(p, q)^{-1}(\sigma_j^2)^{p-1}(1 - \sigma_j^2)^{q-1} \tag{25}$$

with $q=1$ be used to hold σ_j^2 away from zero without restricting its approach to 1. When $m=2$, for example, MAP estimation with this prior adds the penalty function,

$$-\frac{2(p-1)}{d_j^2} \left[\begin{array}{c} a_{j1} \\ a_{j2} \end{array} \right],$$

where

$$d_j^2 = 1 + a_{j1}^2 + a_{j2}^2,$$

to the likelihood equations, and adds the ridge,

$$\frac{2(p-1)}{d_j^4} \left[\begin{array}{cc} d_j^2 - 2a_{j1}^2 & -2a_{j1}a_{j2} \\ -2a_{j1}a_{j2} & d_j^2 - 2a_{j1}^2 \end{array} \right],$$

to the information matrix of the M-step maximum likelihood estimator. We find this method of suppressing Heywood cases to perform well in full-information item factor analysis.

2.5 Computing Times

Computing times depend upon the number of factors, items, subjects, quadrature points, EM cycles, M-step iterations, and the proportion of omitted or not presented items. The preliminary steps of data input and computing starting values are not very time consuming relative to the full-information solution. Most of the time in the latter is accounted for by the evaluation of likelihoods in the E-step; the M-step times are relatively small.

12

Some idea of the overall speed of the present implementation is given by the IBM 3081 cpu time for the test of general science discussed in section 4.3. The total go-step cpu time for a three factor solution with 25 items, 1,178 subjects, $3^3 = 27$ quadrature points, 35 EM cycles, a maximum of five M-step iterations, and numbers of omits as shown in Table 4-6, was 11 minutes and 43 seconds.

3   SIMULATION STUDIES

As a check on both the derivation and the computer implementation, we performed the following analyses of simulated data.

3.1   A One-Factor Test

This simulation demonstrates the capacity of marginal maximum likelihood factor analysis to identify unidimensional item sets in the presence of guessing. To verify that the analysis has no tendency to produce difficulty factors, the item facilities were chosen to span a range larger than is typical of most tests of ability. This was done by setting the item intercepts at equally spaced points between -2.0 and $+2.0$. All item slopes were set at 1.0, corresponding to a factor loading of .707, and all guessing parameters (lower asymptotes) were set at 0.25. Responses with and without guessing were simulated for 1000 subjects drawn randomly from a normal (0,1) distribution of ability.

Three analyses were performed: 1) no guessing assumed in the data or in the analysis; 2) guessing in the data but no guessing assumed in the analysis; 3) guessing assumed in the data and in the analysis. In all of these analyses, the item intercepts and factor loadings were estimated from the data by an EM marginal maximum likelihood solution in which the iterations began from the principal factors of the sample tetrachoric correlation matrix (with communality iteration). Item guessing parameters, on the other hand, were set at their assumed values and not estimated.

It is instructive to examine the effects of guessing and the effect of correction for guessing on the item facilities and the item tetrachoric correlations. These relationships are shown graphically in Figures 3-1 and 3-2. Figure 3-1 confirms the well-known effect of guessing on item facilities. Deviation of the observed facilities from their theoretical values as a function of the true item intercepts from their theoretical values is due entirely to sampling.

Figure 3-2 shows the average tetrachoric correlations for sets of three successive items ordered by facility. When guessing is not assumed or corrected for, the average coefficients are near their theoretical value of .5 at all levels of facility. When guessing is present, but uncorrected, the average tetrachoric coefficients are atten-

13

uated, and the effect becomes greater as the items become harder. At the highest levels of difficulty, most of the correct responses are due to chance successes and the tetrachoric correlation is essentially zero. The effect of this attenuation is to increase the rank of the correlation matrix, and thus to introduce spurious factors in much the same way that variation in item difficulty introduces spurious factors in the analysis of item coefficients. Table 3-1 shows the distinctive pattern of loadings on the spurious second factor that results when guessing effects are ignored in the analysis: items on either extreme of the difficulty continuum tend to have opposite signs.

.

.

.

TABLE 3-1

PRINCIPLE FACTOR LOADINGS FROM SIMULATED DATA WITH GUESSING EFFECT ANALYZED BY NO-GUESSING AND GUESSING MODELS

Item	Non-Guessing Model Principal	Factors	Guessing Model Principal Factor
	1	2	1
1	0.703	0.147	0.761
2	0.719	0.046	0.724
3	0.739	0.215	0.732
4	0.654	−0.029	0.684
5	0.642	0.069	0.660
6	0.689	0.124	0.736
7	0.660	0.065	0.697
8	0.704	0.129	0.755
9	0.580	−0.032	0.697
10	0.561	−0.106	0.697
11	0.574	−0.049	0.710
12	0.583	−0.204	0.765
13	0.505	−0.102	0.715
14	0.393	−0.213	0.665
15	0.407	−0.168	0.704
16	0.329	0.003	0.716
17	0.274	−0.068	0.688
18	0.211	−0.081	0.653
19	0.148	−0.545	0.724
20	0.041	−0.068	0.594
21	0.128	0.069	0.759

*True factor loadings = 0.707
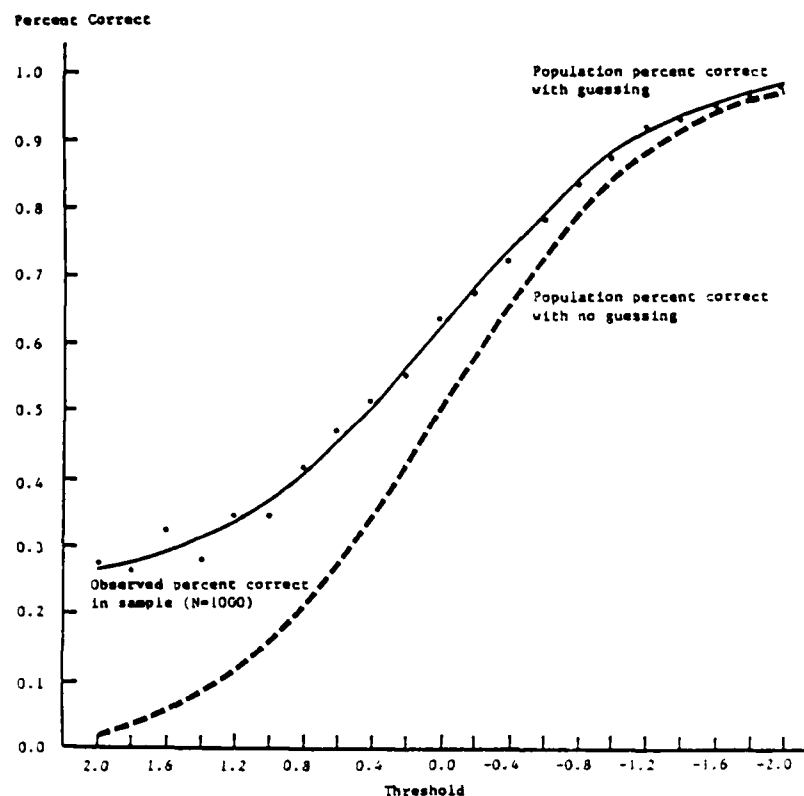
Figure 3-1 Population and sample percent correct as a function of item threshold (Simulated data)

When the guessing model is assumed, both in calculating the tetrachoric correlations and in the response function for the marginal maximum likelihood factor analysis, the deleterious effects of guessing are largely eliminated. As shown in Table 3-2, the likelihood ratio test for the addition of a second factor, which is significant when the no-guessing model was applied to guessing data in Analysis 2, falls to insignificance when guessing is assumed in Analysis 3. The estimated first factor loadings, which were much attenuated in Analysis 2, are raised in Analysis 3 to near their true values.

TABLE 3-2

ANALYSIS OF UNIDIMENSIONAL STIMULATED DATA: CHANGE OF THE
LIKELIHOOD RATIO CHI-SQUARE
UPON ADDING A SECOND FACTOR TO THE MODELS
WITH AND WITHOUT GUESSING

Model	Chi-square	d.f.	p
No Guessing	39.166	20	0.006
Guessing	26.928	20	0.137

These results illustrate the robustness of the analysis in identifying the number of factors and in estimating the factor loadings in the presence of a wide range of item difficulty and of guessing at a typical level of chance success. This relatively successful performance of the method is qualified, however, by its use of assigned rather than estimated guessing parameters. Underestimation of these parameters would certainly leave some effect of guessing in the solution and possibly produce spurious factors. (See also the discussion of the ASVAB Word Knowledge test in Section 4.3.)

3.2   A Two-Factor Test

To demonstrate the power of MML item factor analysis to detect a second factor, a simulation study was conducted based on an analysis of the Auto and Shop Information subtest of the Armed Services Vocational Aptitude Battery (ASVAB). This subtest was constructed from the previously separate Auto Information and Shop Information test items of the earlier Army Classification Battery. As discussed in section 4, three factors were extracted from the observed data for 1,178 cases by a stepwise MML item factor analysis. As shown in Table 3-3, the change in the likelihood ratio chi-square due to inclusion of a second factor was significant, but that due to the third factor was not.
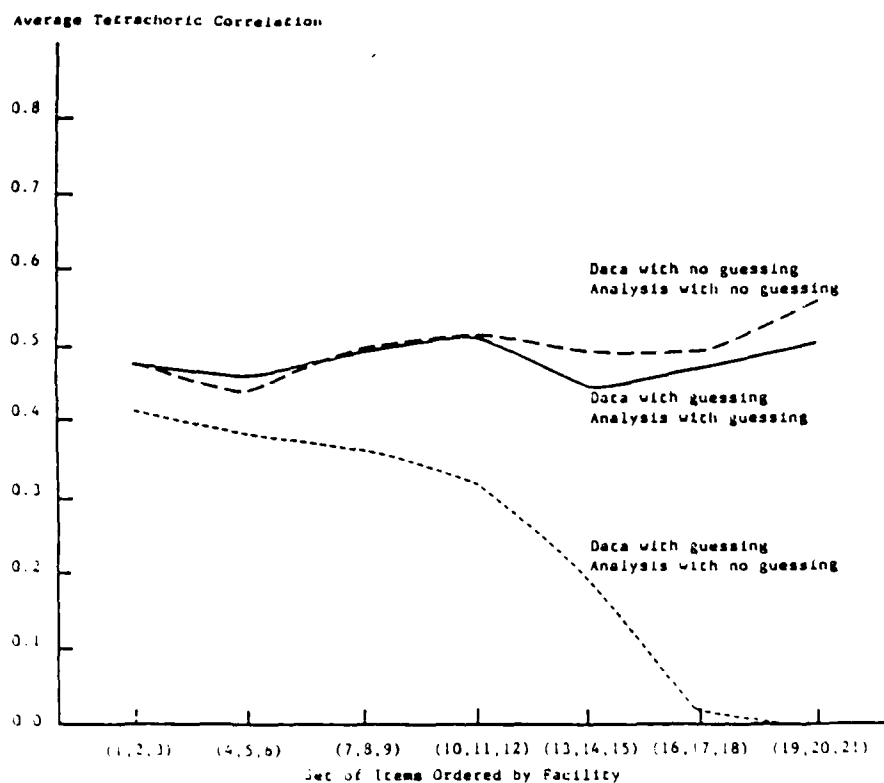
Figure 3-2 Average Tetrachoric Correlations of Sets of Three Successive Items

The resulting estimated factor loadings of the two-factor solution are plotted in the upper panel of Figure 3-3 after orthogonal rotation to the varimax criterion. The axes after oblique rotation to the promax criterion are also shown. Although items 1, 3 and 10, and possibly 2, are misclassified, the plot clearly separates the auto and shop moieties. Based on these loadings for the 25 items, binary scores of 1000 simulated subjects were generated according to the formula (16) with the lower asymptote values shown for the Auto-Shop test in Table 4-5. Factor scores were drawn randomly from a standard normal distribution.

These simulated data were then analyzed by the MML item factor analysis with

TABLE 3-3

CHANGE OF THE LIKELIHOOD RATIO CHI-SQUARE
IN THE ITEM FACTOR ANALYSIS OF THE AUTO
AND SHOP INFORMATION TEST

Factor	Chi-square*	d.f.	p
2 vs. 1	75.6	24	0.000
3 vs. 2	24.7	23	0.363

*Assumed design effect = 2.

lower asymptotes assigned the specified values. Again two significant factors were found. The lower panel of Figure 3-3 gives the resulting varimax rotated factor loadings and promax rotated axes. The MML estimates based on the simulated responses are very similar to their generating values.
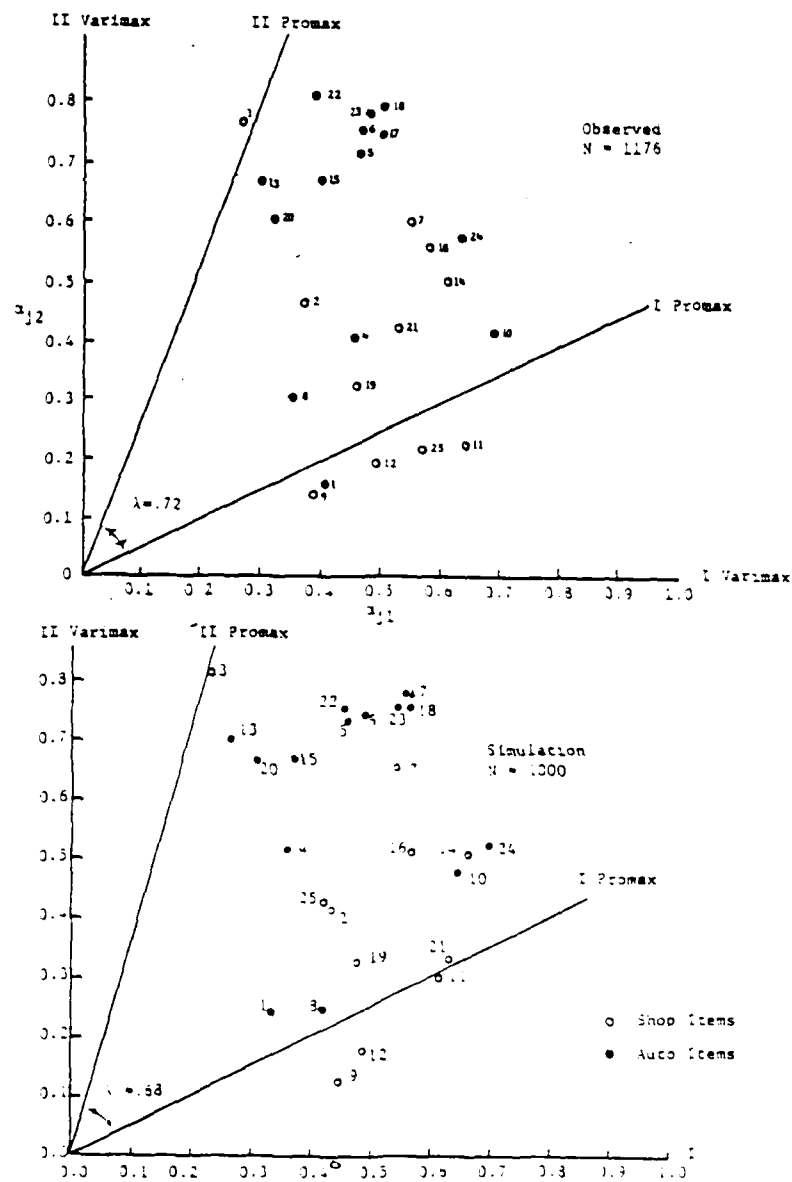
Figure 3-3 Factor Loadings for Observed and Simulated Auto & Shop Information Test

4 APPLICATIONS

In this section, the full-information analysis is applied to a number of empirical data sets.

4.1 Analysis of the LSAT Section 7 With and Without Guessing

Table 4-1 shows the tetrachoric correlations uncorrected and corrected for guessing assuming an asymptote of 0.2 for all items. Note that the correction increases the magnitude of all the coefficients.

TABLE 4-1

TETRACHORIC CORRELATION COEFFICIENTS
OF THE LSAT-7 ITEMS (COEFFICIENTS CORRECTED
FOR GUESSING ABOVE THE DIAGONAL: $g = 0.2$)
($N = 1000$)

		Item				
		1	2	3	4	5
	1	1.000	0.294	0.358	0.401	0.344
	2	0.226	1.000	0.567	0.288	0.174
Item	3	0.291	0.432	1.000	0.376	0.325
	4	0.296	0.204	0.277	1.000	0.214
	5	0.286	0.135	0.265	0.161	1.000

Marginal Log Likelihood
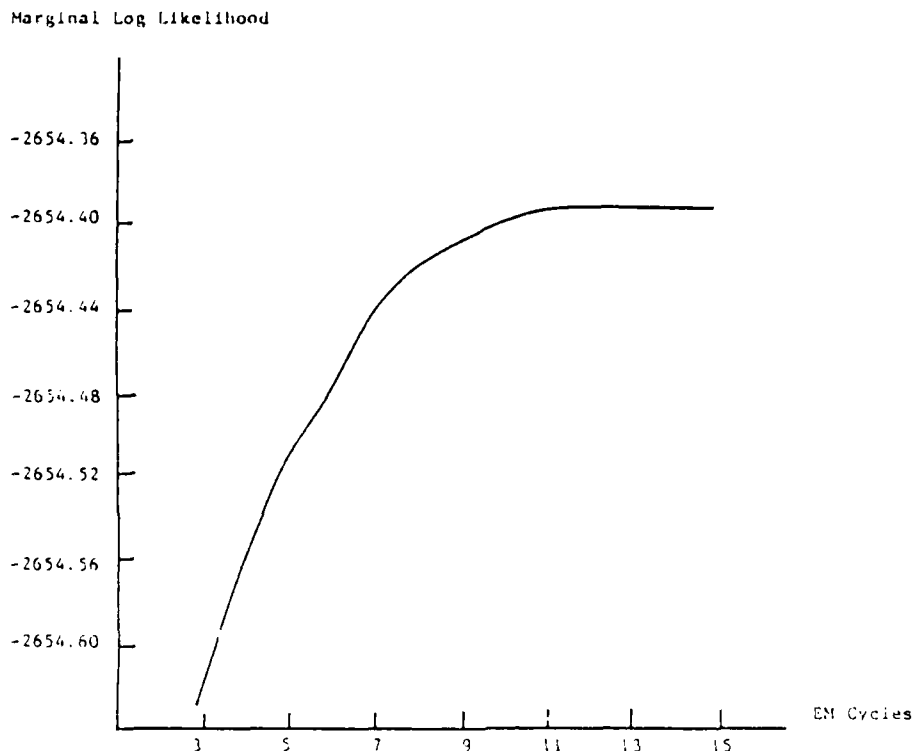
Figure 4-1 Increase in Marginal Log Likelihood in Successive EM Cycles of a Two Factor Solution Without Guessing: LSAT-7

Figure 4-1 shows the increase in marginal log likelihood in successive EM cycles of a two factor solution without guessing. Even with the use of Ramsay accelerator, the likelihood increases slowly as the solution point is approached. Twelve cycles were required for convergence.

TABLE 4-2

CHI-SQUARE STATISTICS FOR THE TWO-FACTOR
STEPWISE ANALYSIS WITH AND WITHOUT GUESSING:
LSAT-7
$(N = 1000)$

	No Guessing			Guessing		
	Chi-square	D.F.	P	Chi-square	D.F.	P
One-Factor	31.66	21	0.063	32.94	21	0.047
Two-Factor	22.86	17	0.154	24.80	17	0.099
Change	8.80	4	0.066	8.14	4	0.086

TABLE 4-3

LSAT-7 RESIDUAL CORRELATIONS
GUESSING ABOVE DIAGONAL

		Item				
		1	2	3	4	5
	1	—	0.016	-0.005	0.043	0.032
	2	0.009	—	0.000	0.005	-0.048
Item	3	-0.024	0.003	—	0.037	0.050
	4	0.026	-0.003	0.018	—	-0.036
	5	0.017	-0.015	0.034	-0.042	—
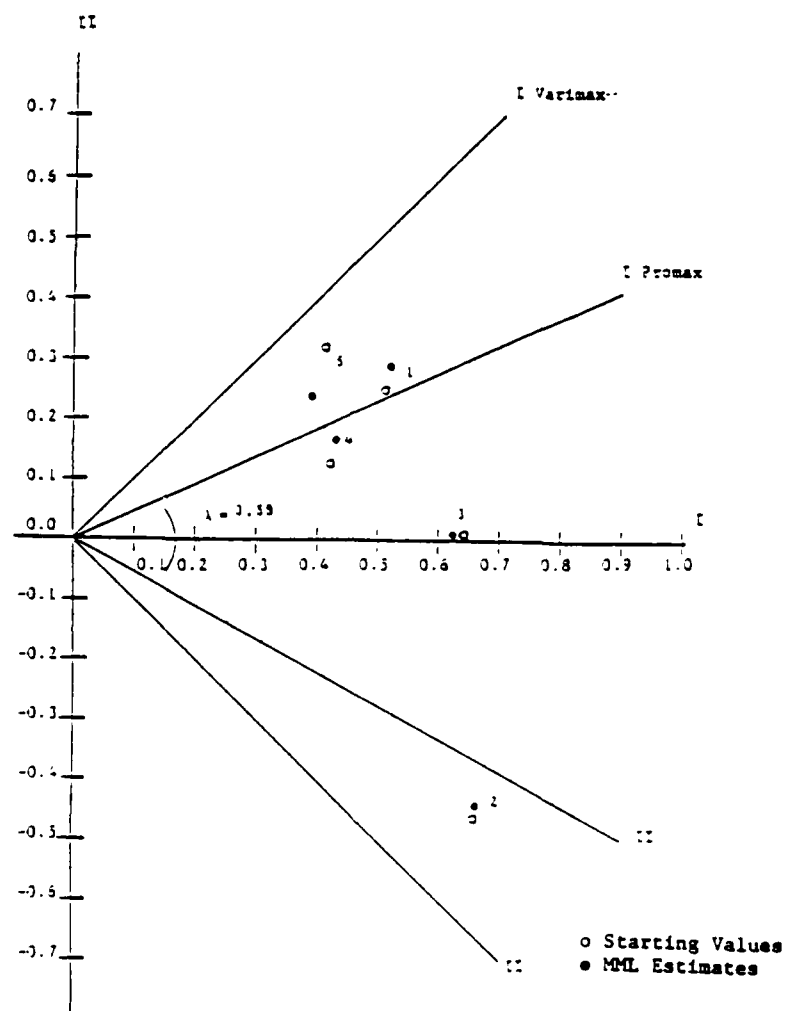
Figure 4-2 Principal Factor Starting Values and MML Estimates of Factor Loadings

With five items and 1000 subjects, these data permit the accurate calculation of goodness-of-fit chi-square as well as change chi-squares, as seen in Table 4-2. Both give evidence of a marginally significant second factor, and there is no indication

24

that the guessing correction improves the solution. Similar conclusions are indicated by the residuals from the tetrachoric coefficients shown in Table 4-3.

Figure 4-2 shows the principal factor starting values (open circles) and MML estimates of the factor loadings from the nonguessing solution (closed circles). It is apparent that loadings on the second factor are changed most by the full-information solution, and that the item with the most extreme correlations, item 5, is most affected. The factor axes rotated to the varimax and promax criteria show that item 2 most clearly determines the second factor.

4.2 The Quality of Life

Campbell, Converse, and Rodgers (1976) assessed 13 aspects of quality of life in 1800 randomly selected respondents to a NORC survey. Respondents rated each quality in terms of their satisfaction with that aspect of their lives. For present purposes, these ratings were dichotomized at the neutral category, and a random sample of 1000 cases was selected. A five factor solution for these data is displayed in Table 4-4. Inspection of Table 4-4 clearly reveals five easily interpretable dimensions underlying the quality of life; 1) health, 2) satisfaction with the living environment (i.e. neighborhood and house quality), 3) satisfaction with everyday life (i.e. job, leisure, friends, family and overall life), 4) financial satisfaction (i.e. savings and standard of living) and 5) satisfaction with self. In terms of level of satisfaction as indicated by the item thresholds in Table 4-4, most respondents were satisfied with their health, family and friends; however, only the most satisfied respondents also reported satisfaction with their savings and education.

As a further verification of the factor solution, a limited-information GLS analysis was also performed (Muthén, 1978). The results of this analysis, employing Muthén's LISCOMP program, are shown in Table 4-4; they correspond closely to those of the full-information solution. Parameter estimates are quite similar and the chi-square statistics for the improvement of fit with the addition of each new factor were virtually identical. The concordance between these two computationally different methods is taken as strong support for the validity of both the methods and the correctness of their implementations.

4.3 Power Tests of the Armed Services Vocational Aptitude Battery ASVAB) Form 8A

The latent dimensionality of each of the eight power tests of the Armed Services Vocational Aptitude Battery (ASVAB) was examined in a ten-percent random sample of data from the Profile of American Youth Study (see Bock and Moore, 1986).

TABLE 4-4

QUALITY OF LIFE DATA: ANALYSIS BY MARGINAL MAXIMUM LIKELIHOOD AND GENERALIZED LEAST SQUARES

Satisfaction With	Thresh-old	Health		Living Envirnmt.		Finance		Everyday Life		Self	
		MML	GLS	MML	GLS	MML	GLS	MML	GLS	MML	GLS
1. Neighborhood	0.83	0.15	0.10	0.63	0.65	0.11	0.08	0.29	0.34	0.05	0.07
2. Education	0.02	0.24	0.12	0.18	0.19	0.24	0.22	0.11	0.20	0.29	0.29
3. Job	0.81	0.08	−0.03	0.19	0.18	0.33	0.32	0.57	0.62	0.07	0.08
4. Leisure	0.67	0.34	0.22	0.20	0.20	0.18	0.16	0.42	0.50	0.38	0.40
5. Health	1.07	0.64	1.10*	0.04	0.07	0.09	0.09	0.12	0.18	0.26	0.20
6. Standard of Living	0.47	0.07	0.02	0.30	0.34	0.64	0.57	0.35	0.38	0.23	0.24
7. Friends	0.98	0.10	0.04	0.19	0.18	0.12	0.08	0.49	0.50	0.28	0.33
8. Savings	−0.17	0.17	0.11	0.18	0.16	0.72	0.83	0.20	0.20	0.13	0.17
9. House	0.70	0.02	0.01	0.73	0.75	0.29	0.28	0.13	0.13	0.18	0.19
10. Family	0.95	0.19	0.10	0.15	0.15	0.15	0.16	0.60	0.61	0.22	0.25
11. Life	0.89	0.31	0.18	0.20	0.23	0.34	0.29	0.56	0.66	0.41	0.41
12. Life in U.S.	0.85	0.43	0.25	0.14	0.17	0.12	0.11	0.40	0.48	0.02	0.10
13. Self	0.90	0.30	0.15	0.11	0.11	0.21	0.20	0.35	0.32	0.74	0.86
Change of χ^2				66.8	64.2	41.4	45.8	33.2	28.0	21.3	17.4
D.F.				12	12	11	11	10	10	9	9

*Heywood case

The data base from which this sample was extracted consisted of ASVAB item responses of 11,817 members of the Youth Panel of the National Longitudinal Study of Labor Force Participation (NLS). The battery was administered under standard conditions by personnel of the National Opinion Research Center (NORC). Because the panel members were selected in a clustered probability sample, the design effect is greater than unity, and some adjustment of the conventional random sampling statistical criteria is necessary.

Previous analysis of these data by Bock and Mislevy (1981) provides the estimates of the lower asymptote parameters for each item shown in Table 4-5. These values were used when the guessing model was assumed in the full-information item factor analyses. Inasmuch as the examinees were given no explicit instructions about guessing or omitting items, it seemed appropriate to score omits as incorrect. Ei-

ther because they run out of time or find the items too difficult, however, some examinees stop responding before they complete all the items on a given test. In these cases, we considered all items following the last non-omitted item to be "not presented". This avoids the spurious association among items later in the test when it is not operating strictly as a power test for all examinees. (See, however, the special handling of the Word Knowledge test.)

TABLE 4-5
ASVAB 8A GUESSING PARAMETER VALUES FROM BOCK & MISLEVY (1981)*

Item	GS	AR	WK	PC	AS	MK	MC	EI
1	0.204	0.210	0.202	0.296	0.221	0.197	0.218	0.197
2	0.213	0.202	0.191	0.203	0.207	0.179	0.232	0.363
3	0.220	0.149	0.217	0.252	0.204	0.135	0.194	0.209
4	0.226	0.173	0.190	0.242	0.228	0.290	0.198	0.154
5	0.159	0.230	0.163	0.127	0.220	0.181	0.137	0.208
6	0.174	0.148	0.249	0.308	0.175	0.178	0.477	0.171
7	0.291	0.207	0.229	0.201	0.255	0.321	0.334	0.262
8	0.185	0.183	0.302	0.196	0.194	0.139	0.139	0.171
9	0.204	0.160	0.161	0.188	0.189	0.305	0.178	0.264
10	0.189	0.250	0.189	0.152	0.215	0.225	0.126	0.121
11	0.218	0.096	0.207	0.196	0.253	0.309	0.226	0.277
12	0.374	0.160	0.157	0.228	0.174	0.133	0.264	0.192
13	0.262	0.261	0.329	0.186	0.135	0.198	0.300	0.147
14	0.188	0.139	0.217	0.191	0.254	0.159	0.227	0.181
15	0.248	0.182	0.151	0.186	0.195	0.234	0.150	0.150
16	0.145	0.340	0.233		0.195	0.309	0.180	0.186
17	0.230	0.289	0.173		0.196	0.119	0.325	0.234
18	0.197	0.311	0.257		0.218	0.211	0.257	0.121
19	0.290	0.200	0.266		0.130	0.295	0.342	0.200
20	0.162	0.079	0.194		0.222	0.262	0.211	0.166
21	0.156	0.185	0.334		0.225	0.280	0.255	
22	0.230	0.250	0.275		0.178	0.180	0.275	
23	0.207	0.262	0.250		0.063	0.120	0.124	
24	0.128	0.205	0.335		0.159	0.127	0.189	
25	0.272	0.219	0.320		0.196	0.152	0.167	
26		0.152	0.200					
27		0.095	0.194					
28		0.152	0.267					
29		0.139	0.287					
30		0.203	0.275					
31			0.264					
32			0.300					
33			0.276					
34			0.168					
35			0.159					

*See text for test names.

28

The results of the item factor analyses, with the estimated factor loadings shown both in their principal factor and promax rotations, are shown in Tables 4-6 through 4-13. These tables include the change chi-squares, degrees of freedom, and probability levels due to inclusion of additional factors. Also shown are percents of variance associated with each of the principal factors (i.e., the percent that the corresponding latent root of the reproduced item-correlation matrix is of the trace of that matrix) and the intercorrelations of the promax factors.

TABLE 4-6
ASVAB 8A
ITEM FACILITIES, ATTEMPTS, STANDARD DIFFICULTIES, AND FACTOR LOADINGS
GENERAL SCIENCE ($N = 1,178$)

Item	Facility	Attempts	Diffi-culty	Principal Factors 1	Principal Factors 2	Promax Factors 1	Promax Factors 2
1	0.843	1177	−0.827	0.710	−0.319	0.008	0.773
2	0.758	1177	−0.463	0.737	0.092	0.581	0.201
3	0.726	1176	−0.327	0.794	0.105	0.633	0.209
4	0.669	1176	−0.141	0.626	0.153	0.595	0.064
5	0.722	1176	−0.392	0.628	0.141	0.580	0.083
6	0.765	1176	−0.513	0.779	−0.264	0.126	0.724
7	0.672	1176	−0.024	0.675	−0.235	0.101	0.637
8	0.805	1176	−0.678	0.548	−0.234	0.023	0.579
9	0.726	1176	−0.354	0.711	0.093	0.566	0.188
10	0.709	1176	−0.322	0.590	0.157	0.578	0.043
11	0.662	1176	−0.120	0.715	−0.036	0.394	0.374
12	0.513	1176	0.835	0.719	0.319	0.876	−0.128
13	0.472	1175	0.633	0.884	0.242	0.875	0.055
14	0.608	1174	0.011	0.620	0.069	0.478	0.181
15	0.685	1171	−0.164	0.542	−0.138	0.149	0.440
16	0.638	1167	−0.139	0.609	−0.294	−0.021	0.691
17	0.618	1163	0.052	0.566	−0.484	−0.304	0.941
18	0.384	1155	0.795	0.900	0.045	0.618	0.342
19	0.473	1150	0.778	0.765	−0.102	0.336	0.489
20	0.477	1142	0.390	0.628	−0.095	0.261	0.417
21	0.353	1131	0.844	0.651	0.176	0.642	0.043
22	0.343	1125	1.121	0.799	−0.087	0.377	0.484
23	0.338	1104	1.113	0.701	0.389	0.960	−0.235
24	0.215	1091	1.365	0.891	0.034	0.598	0.353
25	0.358	1055	1.270	0.933	0.044	0.637	0.358

Adding Factor	Chi-Square* Change	D.F.	P	Percent of Variance		Factor Correlations		
2	67.2	24	0.000	51.457	4.391	1	1.000	
3	14.2	23	0.922			2	0.740	1.000

*Assumed design effect = 2.

30

TABLE 4-7
ASVAB 8A
ITEM FACILITIES, ATTEMPTS, STANDARD DIFFICULTIES, AND FACTOR LOADINGS
ARITHMETIC REASONING ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors 1	Principal Factors 2	Promax Factors 1	Promax Factors 2
1	0.896	1176	−1.096	0.480	0.226	0.042	0.497
2	0.896	1176	−1.109	0.628	0.448	−0.164	0.894
3	0.703	1176	−0.335	0.787	−0.118	0.767	0.035
4	0.662	1176	−0.158	0.842	−0.064	0.732	0.138
5	0.606	1176	0.087	0.746	−0.021	0.598	0.178
6	0.665	1176	−0.222	0.728	−0.042	0.614	0.141
7	0.745	1176	−0.366	0.521	0.171	0.151	0.422
8	0.680	1176	−0.215	0.702	−0.074	0.639	0.083
9	0.645	1176	0.126	0.748	−0.158	0.795	−0.039
10	0.606	1176	0.128	0.893	0.119	0.508	0.444
11	0.551	1176	0.067	0.876	−0.083	0.785	0.117
12	0.526	1175	0.219	0.768	−0.075	0.692	0.099
13	0.560	1175	0.321	0.773	0.034	0.539	0.275
14	0.501	1175	0.284	0.821	−0.209	0.923	−0.099
15	0.571	1170	0.151	0.818	−0.183	0.891	−0.067
16	0.565	1167	0.578	0.839	−0.046	0.705	0.165
17	0.478	1167	0.774	0.849	−0.163	0.879	−0.019
18	0.459	1166	0.886	0.908	0.038	0.636	0.319
19	0.493	1164	0.449	0.722	0.022	0.518	0.240
20	0.308	1162	0.789	0.789	−0.003	0.604	0.220
21	0.386	1159	0.841	0.880	0.004	0.663	0.257
22	0.485	1151	0.640	0.880	−0.110	0.826	0.076
23	0.481	1145	0.616	0.751	0.441	−0.061	0.918
24	0.424	1140	0.763	0.871	0.226	0.339	0.608
25	0.408	1135	0.812	0.878	−0.007	0.677	0.239
26	0.407	1121	0.621	0.744	−0.034	0.615	0.157
27	0.337	1107	0.705	0.793	−0.144	0.809	−0.004
28	0.291	1074	1.122	0.868	0.092	0.527	0.394
29	0.277	1049	1.148	0.820	0.073	0.519	0.350
30	0.392	1018	0.816	0.802	−0.118	0.779	0.040

Adding	Chi-Square* Change	D.F.	P	Percent of Variance		Factor Correlations	
2	93.519	29	0.000	62.469	2.587	1 1.000	
3	27.525	28	0.490			2 0.787	1.000

*Assumed design effect = 2.

31

TABLE 4-8
ASVAB 8A
ITEM FACILITIES, ATTEMPTS, STANDARD DIFFICULTIES, AND FACTOR LOADINGS
WORD KNOWLEDGE ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors		Promax Factors	
				1	2	1	2
1	0.914	1176	−1.173	0.708	0.158	0.141	0.606
2	0.902	1176	−1.118	0.725	0.153	0.158	0.606
3	0.857	1176	−0.828	0.795	0.146	0.209	0.628
4	0.870	1176	−0.908	0.576	0.223	−0.040	0.650
5	0.882	1176	−0.997	0.709	0.358	−0.186	0.938
6	0.812	1176	−0.497	0.917	0.148	0.273	0.693
7	0.834	1176	−0.733	0.677	−0.070	0.494	0.215
8	0.797	1176	−0.466	0.891	−0.094	0.655	0.279
9	0.621	1176	−0.040	0.717	0.077	0.277	0.478
10	0.866	1175	−0.889	0.817	0.131	0.245	0.615
11	0.726	1174	−0.302	0.806	0.042	0.385	0.462
12	0.787	1174	−0.578	0.702	0.247	−0.008	0.751
13	0.806	1174	−0.420	0.872	0.099	0.329	0.588
14	0.678	1173	−0.078	0.880	0.234	0.112	0.817
15	0.717	1171	−0.322	0.843	−0.055	0.563	0.320
16	0.761	1170	−0.380	0.788	0.055	0.354	0.475
17	0.672	1169	−0.077	0.931	0.251	0.113	0.870
18	0.723	1165	−0.226	0.792	−0.175	0.731	0.096
19	0.635	1161	0.100	0.781	−0.240	0.830	−0.016
20	0.752	1160	−0.368	0.831	0.059	0.370	0.503
21	0.723	1158	−0.090	0.807	−0.152	0.700	0.143
22	0.624	1152	0.217	0.934	−0.015	0.550	0.430
23	0.560	1146	0.319	0.850	−0.276	0.928	−0.043
24	0.530	1141	0.672	0.786	−0.238	0.830	−0.011
25	0.547	1132	0.523	0.845	0.022	0.439	0.448
26	0.581	1121	0.243	0.895	-0.033	0.557	0.382
27	0.551	1110	0.303	0.760	-0.098	0.587	0.209
28	0.657	1098	0.084	0.723	−0.143	0.638	0.117
29	0.486	1083	0.756	0.808	−0.103	0.621	0.224
30	0.517	1065	0.588	0.732	−0.222	0.773	-0.010
31	0.834	1050	-0.402	0.845	0.037	0.415	0.473
32	0.473	1036	0.862	0.706	-0.192	0.710	0.027
33	0.478	1017	0.873	0.908	-0.158	0.767	0.182
34	0.561	1003	0.348	0.811	0.038	0.393	0.458
35	0.509	985	0.504	0.878	-0.147	0.733	0.185

Adding Factor	Chi-Square*	D.F.	P	Percent of Variance		Factor Correlations		
Factor	Change							
2	111.470	34	0.000	64.863	2.650	1	1.000	
						2	0.815	1.000

* Assumed design effect = 2.

TABLE 4-9
ASVAB 8A
ITEM FACILITIES, ATTEMPTS, STANDARD DIFFICULTIES, AND FACTOR
LOADINGS
PARAGRAPH COMPREHENSION ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factor 1
1	0.747	1176	−0.322	0.786
2	0.841	1176	−0.832	0.676
3	0.772	1176	−0.486	0.899
4	0.685	1175	−0.189	0.610
5	0.658	1174	−0.247	0.800
6	0.670	1173	−0.024	0.731
7	0.658	1173	−0.162	0.663
8	0.733	1173	−0.422	0.574
9	0.712	1169	−0.350	0.747
10	0.478	1166	0.321	0.735
11	0.723	1160	−0.382	0.483
12	0.566	1150	0.183	0.711
13	0.735	1136	−0.402	0.761
14	0.609	1102	0.008	0.698
15	0.505	1085	0.283	0.143
Adding Factor	Chi-square* Change	D.F.	P	Percent of Variance
2	11.6	14	0.640	47.497

*Assumed design effect $= 2$.

33

TABLE 4-10

ASVAB 8A
STANDARD DIFFICULTIES, AND FACTOR LOADINGS
AUTO AND SHOP INFORMATION ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors		Promax Factors	
				1	2	1	2
1	0.704	1176	−0.300	0.381	0.201	−0.058	0.471
2	0.768	1176	−0.565	0.592	−0.026	0.368	0.268
3	0.602	1176	0.015	0.753	−0.295	0.822	−0.019
4	0.799	1176	−0.651	0.604	0.079	0.233	0.417
5	0.615	1176	−0.038	0.849	−0.117	0.635	0.275
6	0.491	1176	0.263	0.876	−0.132	0.671	0.268
7	0.467	1176	0.532	0.818	0.025	0.426	0.454
8	0.603	1176	−0.037	0.465	0.168	0.034	0.469
9	0.633	1176	−0.104	0.356	0.200	−0.070	0.457
10	0.545	1176	0.188	0.762	0.248	0.093	0.730
11	0.551	1175	0.265	0.584	0.339	−0.130	0.764
12	0.556	1174	0.093	0.469	0.242	−0.063	0.572
13	0.558	1174	0.006	0.701	−0.210	0.678	0.071
14	0.582	1174	0.131	0.779	0.127	0.267	0.573
15	0.469	1171	0.390	0.769	−0.137	0.617	0.206
16	0.467	1166	0.412	0.806	0.081	0.344	0.524
17	0.379	1161	0.710	0.895	−0.105	0.644	C.316
18	0.383	1157	0.791	0.930	−0.137	0.708	0.289
19	0.593	1154	−0.092	0.545	0.138	0.120	0.469
20	0.477	1147	0.447	0.666	−0.149	0.576	0.137
21	0.379	1132	0.875	0.655	0.123	0.202	0.505
22	0.379	1126	0.697	0.870	−0.237	0.809	0.121
23	0.262	1114	0.802	0.906	−0.143	0.703	0.268
24	0.273	1093	1.086	0.841	0.111	0.323	0.583
25	0.371	1075	0.780	0.536	0.286	−0.085	0.667

Adding Factor	Chi-Square* Change	D.F.	P	Percent of Variance		Factor Correlations		
2	75.6	24	0.000	51.272	3.243	1	1.000	
3	24.7	23	0.363			2	0.731	1.000

*Assumed design effect = 2.

TABLE 4-11

ASVAB 8A

ITEM FACILITIES, STANDARD DIFFICULTIES, AND FACTOR LOADINGS

MATHEMATICAL KNOWLEDGE ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors		Promax Factors	
				1	2	1	2
1	0.803	1175	−0.692	0.780	−0.376	−0.230	1.054
2	0.721	1174	−0.414	0.620	−0.163	0.068	0.582
3	0.535	1174	0.081	0.768	−0.081	0.306	0.494
4	0.652	1174	0.035	0.845	−0.023	0.460	0.418
5	0.680	1174	−0.280	0.576	−0.128	0.106	0.497
6	0.519	1173	0.232	0.843	0.013	0.524	0.350
7	0.608	1173	0.212	0.922	0.151	0.824	0.126
8	0.523	1173	0.150	0.684	−0.007	0.393	0.317
9	0.598	1173	0.222	0.836	0.131	0.736	0.125
10	0.561	1173	0.180	0.746	0.006	0.454	0.320
11	0.509	1171	0.599	0.780	0.078	0.606	0.200
12	0.422	1170	0.434	0.839	−0.179	0.168	0.710
13	0.469	1168	0.422	0.945	0.024	0.605	0.374
14	0.386	1166	0.626	0.907	−0.046	0.452	0.490
15	0.388	1163	0.902	0.597	−0.051	0.259	0.362
16	0.493	1159	0.689	0.946	0.080	0.708	0.269
17	0.379	1158	0.578	0.889	0.111	0.732	0.185
18	0.431	1158	0.624	0.934	0.122	0.784	0.190
19	0.502	1157	0.601	0.834	−0.253	0.029	0.846
20	0.419	1152	0.813	0.854	0.065	0.627	0.257
21	0.375	1147	1.046	0.884	0.199	0.891	0.018
22	0.318	1143	0.991	0.940	0.147	0.828	0.141
23	0.269	1135	1.049	0.905	−0.067	0.414	0.527
24	0.264	1114	1.027	0.778	−0.116	0.247	0.565
25	0.281	1084	1.023	0.923	0.144	0.813	0.139

Adding Factor	Chi-Square* Factor	D.F. Change	P	Percent of Variance		Factor Correlations		
2	77.7	24	0.000	68.954	1.903	1	1.000	
3	28.0	23	0.216			2	0.856	1.000

*Assumed design effect $= 2$.

35

TABLE 4-12

ASVAB 8A
ITEM FACILITIES, ATTEMPTS, STANDARD DIFFICULTIES, AND FACTOR LOADINGS
MECHANICAL COMPREHENSION ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors 1	2	Promax Factors 1	2
1	0.865	1175	−0.948	0.496	0.032	0.378	0.144
2	0.727	1175	−0.373	0.744	−0.076	0.729	0.024
3	0.740	1175	−0.464	0.587	−0.015	0.516	0.088
4	0.380	1175	0.757	0.748	0.028	0.597	0.186
5	0.543	1175	0.079	0.692	−0.118	0.740	−0.052
6	0.580	1174	0.860	0.766	−0.017	0.671	0.120
7	0.518	1174	0.622	0.864	−0.011	0.746	0.147
8	0.557	1174	0.042	0.792	0.010	0.658	0.166
9	0.617	1174	−0.081	0.519	−0.151	0.636	−0.134
10	0.530	1174	0.096	0.832	0.237	0.395	0.524
11	0.609	1174	0.030	0.427	−0.077	0.462	−0.038
12	0.512	1174	0.418	0.814	−0.078	0.791	0.034
13	0.598	1174	0.196	0.779	0.072	0.754	0.037
14	0.518	1173	0.258	0.753	0.607	−0.155	1.081
15	0.498	1173	0.237	0.712	0.031	0.562	0.184
16	0.541	1170	0.157	0.555	−0.145	0.660	−0.118
17	0.472	1168	0.827	0.800	−0.212	0.954	−0.175
18	0.446	1163	0.702	0.868	0.182	0.498	0.445
19	0.436	1157	1.292	0.847	−0.003	0.722	0.156
20	0.474	1146	0.461	0.639	−0.042	0.596	0.057
21	0.397	1138	0.905	0.830	−0.169	0.923	−0.103
22	0.381	1124	1.107	0.786	0.068	0.577	0.254
23	0.330	1100	0.750	0.725	−0.010	0.627	0.123
24	0.386	1078	0.718	0.696	−0.057	0.664	0.044
25	0.327	1062	0.891	0.797	−0.083	0.783	0.024

Adding	Chi-Square* Change	D.F.	P	Percent of Variance		Factor Correlations	
2	30.0	24	0.185	53.643	2.527	1 1.000	
3	16.0	23	0.858			2 0.766	1.000

*Assumed design effect = 2.

36

TABLE 4-13

ASVAB 8A
ITEM FACILITIES, STANDARD DIFFICULTIES, AND FACTOR LOADINGS
ELECTRONICS INFORMATION ($N = 1,178$)

Item	Facility	Attempts	Difficulty	Principal Factors 1
1	0.757	1176	−0.512	0.619
2	0.674	1176	0.059	0.846
3	0.639	1176	−0.105	0.761
4	0.662	1176	−0.241	0.761
5	0.703	1176	−0.304	0.607
6	0.625	1176	−0.094	0.764
7	0.636	1175	−0.007	0.699
8	0.605	1174	−0.048	0.676
9	0.652	1173	−0.057	0.564
10	0.496	1173	0.200	0.682
11	0.415	1171	0.927	0.628
12	0.420	1169	0.621	0.814
13	0.376	1164	0.643	0.724
14	0.458	1161	0.443	0.387
15	0.403	1157	0.582	0.805
16	0.394	1150	0.707	0.670
17	0.252	1138	1.942	0.704
18	0.389	1131	0.542	0.611
19	0.405	1115	0.714	0.731
20	0.289	1101	1.153	0.780

Adding Factor	Chi-Square* Change	D.F.	P	Percent of Variance
2	21.8	19	0.296	48.879

*Assumed design effect $= 2$.

Except in one instance discussed below, the factors found by the full-information analysis to be statistically significant corresponded to obvious and often cognitively interesting features of the items. Although we cannot exhibit actual items from this test, which is still secure, we can convey descriptively the nature of the factors. Those readers who have access to the test can check our interpretation by examining the items in connection with the factor loadings in the tables. The promax loadings are most useful for this purpose. The number of EM cycles was 35 in each case.

General Science (GS) (Table 4-6). Even with the guessing accounted for, a

significant second factor is found. The corresponding change in chi-square is more than five times its degrees of freedom and would remain significant with an assumed design effect as large as 2.0. The promax factors are easily interpreted. The first is essentially physical science, and the second is biological science—or more precisely, health science. These factors are substantially correlated ($r = 0.740$), reflecting the large percent of variance (51.5) attributable to the first principal factor in contrast with 4.4 percent for the second.

The finding of two factors in GS agrees with the observation of Bock and Mislevy (1981) that there is an item-by-sex interaction in this test such that male examinees tend to do better on physical science items and female examinees better on biological and health science items. These results, in addition to the fact that various civilian and military occupational specialties divide along the same lines, suggest the desirability of scoring the physical science and biological science content of the General Science test separately.

Arithmetic Reasoning (AR) (Table 4-7). There is clear evidence for a significant second factor in this test, but not for a third factor if a design effect of 2.0 is assumed. The second factor makes a very minor contribution to variance, however, and is represented by only three items with high promax loadings. These items involve computation of interest, suggesting some sort of business arithmetic factor. Although additional items might be added to better define such a factor, it appears to be of minor importance in assessing arithmetic reasoning ability.

Word Knowledge (WK) (Table 4-8). More strongly than other tests in the ASVAB, Word Knowledge appears in Form 8 with its items ordered from easy to hard in difficulty. It also has a relatively short time limit—11 minutes for 35 items. As a consequence of these two conditions, the question of how to handle omitted responses at the end of the test is troublesome. Omitted items could mean either that the examinee left off answering because the words became too difficult, or that he ran out of time. If we assume the former, then the omitted responses should be considered incorrect and assigned the guessing probability of a correct response so as to be more equivalent to non-omitted responses earlier in the test. If we assume the latter, the omitted responses following the last non-omitted responses should be treated as not presented.

Considering that the frequency of omitted responses at the end of the WK test is relatively high (see Table 4-8), and assuming that the prescribed time limit had been adequately pretested, we have concluded that, for purposes of the item factor analysis, the omitted items should be assigned the guessing probability of success for that item rather than treated as not presented to that examinee. Scored in this way, WK shows clear evidence of a second significant factor (Table 4-8).

The interpretation of this factor is, however, not at all obvious. The principal

38

factor pattern in Table 4-8 bears no apparent relationship to the item content, but resembles instead the pattern for a "difficulty" factor encountered when phi coefficients are analyzed, or the pattern found in section 3.1 when guessing effects were ignored. That is, the loadings of the second principal factor tend, with only a few exceptions, to be opposite in sign for easy and hard items. Similarly, the promax factors, which are highly correlated, divide the items with respect to difficulty or, equivalently, ordinal position in the test.

Attributing the significant second factor to effects of difficulty or guessing would seem to be ruled out, however, by our demonstration in the simulation study of section 3.1 that the present solution is free of these artifacts. To eliminate the possibility that the solution is influenced by our decision to score not reached items as omitted, we performed an additional analysis treating these items as not presented; again, a significant second factor appeared.

We have found similar evidence for two factors in an analysis of Word Knowledge items constructed for the projected computerized adaptive version of the ASVAB tests. Data from approximately 2600 recruits responding to paper and pencil versions of 47 of these items showed a change chi-square of 454.3 on 46 degrees of freedom, upon addition of a second factor. The percent of principal factor variance attributable to these factors was 59.95 and 1.70, respectively, only four or five items showed appreciable loading on the second factor.

There was no clear relationship between the second factor loadings and the item difficulties. Nor was there any interpretable relationship of the loadings to the content of the item stems or alternatives.

As a further test of the validity of these results; we performed the same analysis on simulations of responses to these items carried out by Charles Davis (Davis, 1986) using non-parametric item response functions fitted to the above ASVAB data by Michael Levine (Levine, 1986). We used the BILOG program of Mislevy and Bock (1984) to estimate the guessing parameters required for the item factor analysis. These simulated data, which were of course strictly unidimensional, showed no decrease in chi-square upon introduction of a second factor.

We construe these results as evidence that the second factor in the Word Knowledge item responses is non-antifactual, but that it corresponds to similarities between items that are not easy to associate with item content. Word Knowledge items have little in the way of systematic features that might suggest a cognitive basis for classifying the items. Until such features can be identified, the minor departures from conditional independence arising from them will have to be ignored when estimating a unidimensional ability in word knowledge.

Paragraph Comprehension (PC) (Table 4-9). Only one factor was found. We had thought that the several paragraphs on which these items are based would appear

as factors, but this was not the case. There is no evidence of failure of conditional independence in this test. Items 11 and 15 have rather poor discriminating power.

Auto and Shop Information (AS) (Table 4-10). This test, composed of items based on the Auto Information and Shop Information tests of the earlier Army Classification Battery, exhibited a significant and very clear two-factor pattern separating the two types of items as already shown in Figure 3-3. As mentioned in section 3.2, the pattern indicates that a few of the items are misclassified. Although a third factor could be extracted in which a few of the loadings suggested a distinction between wood-shop and metal-shop items, it was not significant when a design effect of 2.0 was assumed and is not reported here.

Mathematics Knowledge (MK) (Table 4-11). Two factors of mathematics knowledge are statistically significant; the third is not when a design effect of 2.0 is assumed. Items with large loadings on the first promax factor all require knowledge of formal algebra, while those loadings on the second factor involve numerical calculation and mathematical reasoning. If a third factor is extracted (not shown), it tends to separate calculation from reasoning but not clearly so.

Mechanical Comprehension (MC) (Table 4-12). There is perhaps marginal evidence of a second factor in this test, but it is represented by only two items (10 and 14). These items ask about the speed with which something turns, whereas most of the other items ask only about direction of movement or rotation. Item 18, which asks about both direction and speed, loads on both factors. The same is true of item 22, but it loads more on the first factor. The distinction is of minor importance at best.

Electronics Information (EI) (Table 4-13). This test shows no evidence of a significant second factor when a design effect of 2.0 is assumed. Except for number 14, the items are highly uniform in discriminating power.

4.4 DAT Spatial Reasoning

In a study of item features requiring spatial visualizing ability, Zimowski (1985) carried out a full-information item factor analysis of the Spatial Visualization subtest of the current edition of the Differential Aptitude Test battery (Bennett, Seashore, and Wesman; 1974). Examinees were 390 high school seniors from a suburban Chicago school system. The analysis revealed four statistically significant factors. Considering that the test consists exclusively of pattern folding items, we found this result surprising. Upon examining the items loading most heavily on a given factor, we found that they were constructed from basically the same stimulus pattern, but modified with additional marks and features so as to serve as distinct items. Probably the items were drawn in this way to reduce the amount of original art

40

work required.

That these factors could represent distinct cognitive processes seems unlikely. A more plausible explanation is that a correct response on the first encounter with one of these items increases the probability of a correct response to later items from the similar set, while an incorrect response on the first encounter does not lead to an increase. These failures of conditional independence would produce increased associations among items that would appear as a factor. It may be possible to distinguish this type of factor from a genuine cognitive process factor by position effects. Positively associated items should become less difficult as they are preceded by more items from the same dependent set. This sort of violation of standard item-response theoretic assumptions could easily be corrected by avoiding repeated use of similar features among items in the same scale. Unfortunately, this strategy would rule out unidimensional scales consisting of items generated by varying components of a facet design on the item content or formats. This finding is discussed in greater detail in Zimowski (1986).

5   DISCUSSION AND CONCLUSION

Implementation of item factor analysis by marginal maximum likelihood estimation overcomes many of the problems that attend factor analysis of tetrachoric correlation coefficients: it avoids the problem of indeterminate tetrachoric coefficients of extremely easy or difficult items; it readily accomodates effects of guessing, and omitted or not reached items; and it provides a likelihood ratio test of the statistical significance of additional factors. Although the numerical integration used in the MML approach involves heavy computation and limits the procedure to five factors, the number of items that can be analyzed is sufficiently large (up to 100) to make the method useful in practical test development.

The applications of the procedure reported in the present paper show that, in moderately large samples (500 to 1000 cases), minor factors determined by relatively few items can be detected as significant. The sensitivity of the MML method recommends it as an exploratory technique in searching for item features that are responsible for individual differences in cognitive test performance. By the same token, format attributes that may be implicated in failures of conditional independence are easily detected.

The examples presented in section 4.3 suggest that many routinely used tests may contain some items that produce departures from unidimensionality or conditional independence. In many situations such items could be eliminated by including in the same scale only items that are highly homogeneous in all content and format features that are not relevant to the ability dimension in questions. Otherwise,

41

the only practical alternative may be to use a method of computing conditional probabilities of answer patterns that does not assume independence. A procedure for this purpose, based on the so-called "Clark algorithm" for evaluating orthant probabilities of the multivariate normal distribution, has been proposed by Gibbons and Bock (1986).

References

Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1974). *Manual for the Differential Aptitude Tests Forms S and T* (5th edition). New York: The Psychological Corporation.

Bartholomew, D.J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society*, Series B, **42**, 293-321.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, **46**, 443-459.

Bock, R.D., & Moore, E.G.J. (1986). *Advantage and Disadvantage: A Profile of American Youth*. Hillsdale (N.J.): Erlbaum.

Bock, R.D., & Mislevy, R.J. (1981). *Data Quality Analysis of the Armed Services Vocational Aptitude Battery*. Chicago: National Opinion Research Center.

Campbell, A., Converse, P.E. & Rodgers, W.L. (1976). *The Quality of American Life*. New York: Russell Sage Foundation.

Carroll, J.B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, **10**, 1-19.

Carroll, J.B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), *Principles of Modern Psychological Measurement* (pp. 257-282). Hillsdale (N.J.): Erlbaum.

Cristoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, **40**, 5-32.

Davis, C. (1986). Personal Communication.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Series B, **39**, 1-38.

Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, **44**, 169-172.

Gibbons, R.D. & Bock, R.D. (1986). Approximating multivariate normal orthant probabilities. (In preparation.)

Haberman, J.S. (1977). Log-linear models and frequency tables with small expected cells counts. *Annals of Statistics*, **5**, 1148-1169.

Harman, H.H. (1976). *Modern Factor Analysis*, Chicago: The University of Chicago Press.

Hendrickson, A.E., & White, P.O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology*, **17**, 65-70.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443-482.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200.

Lee, S.Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, **46**, 153-160.

Levine, M. (1986). Personal Communication.

Martin, J.K., & McDonald, R.P. (1973). Bayesian estimation in unrestricted factor analysis: a treatment for heywood cases. *Psychometrika*, **40**, 505-517.

Mislevy, R.J. (1984). Personal Communication.

Mislevy, R.J. & Bock, R.D. (1984). BILOG: Maximum Likelihood Item Analysis and Test Scoring—Logistic Model. Mooresville (Ind.): Scientific Software.

Muthén, B. (1978). Contributions to factor analysis of dichotomized variables. *Psychometrika*, **43**, 551-560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categories, and continuous latent variable indicators. *Psychometrika*, **40**, 337-360.

Ramsay, J.O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika*, **40**, 337-360.

Wilson, D., Wood, R. & Gibbons, R.D. (1984). TESTFACT: Test Scoring, Item Statistics, and Item Factor Analysis. Mooresville: (Ind.): Scientific Software.

Zimowski, M.F. (1985). Attributes of Spatial Test Items that Influence Cognitive Processing. Unpublished doctoral dissertation, Department of Behavioral Sciences, University of Chicago, Chicago, IL.

Zimowski, M.F. (1985a). Item factor analysis of ravens progressive matrices and the DAT spatial visualization test. (in preparation)

National Opinion Research Center/Bock NR 475-018

Personnel Analysis Division,
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

AFOSR,
Life Sciences Directorate
Bolling Air Force Base
Washington, DC 20332

Dr. William E. Alley
AFHRL/MOT
Brooks, AFB, TX 78235

Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Special Assistant for Projects,
OASN(M&RA)
5D800, The Pentagon
Washington, DC 20350

Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152

Dr. R. Darrell Bock
University of Chicago
Department of Education
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
NAVTRAEQUIPCEN
Orlando, FL 32813

M.C.S. Jacques Bremond
Centre de Recherches du Service
de Sante des Armees
1 Bis, Rue du Lieutenant Raoul Batany
92141 Clamart, FRANCE

Dr. Robert Brennan
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
NAVOP 01B7
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Director,
Manpower Support
and Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Paul Horst
677 G. Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
NAVOP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Earl Hunt
Department of Psychology
University of Washington
Seattle, WA 98105

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Advanced Statistical
Technologies Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Dept.
501 North Dixon Street
P.O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Patrick Kyllonen
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Anita Lancaster
Accession Policy
OASD/MI&L/MP&FM/AP
Pentagon
Washington, DC 20301

Dr. Daryll Lang
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
The Netherlands

Science and Technology Div.
Library of Congress
Washington, DC 20540

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

Assistant for Evaluation, Analysis,
and MIS, NMPC
N-6C
Washington, DC 20370

Spec. Asst. for Research,
Experimental & Academic Programs,
NTTC (Code 016)
NAS Memphis (75)
Millington, TN 38054

Director,
Research & Analysis Div.,
NAVCRUITCOM Code 22
4015 Wilson Blvd.
Arlington, VA 22203

Dr. David Navon
Institute for Cognitive Science
University of California
La Jolla, CA 92093

Assistant for Long Range
Requirements, CNO Executive Panel
NAVOP OOK
200 N. Beauregard Street Alexandria, VA
22311

Assistant for Planning MANTRAPERS
NAVOP 01B6
Washington, DC 20370

Assistant for MPT Research,
Development and Studies, NAVOP 01B7
Washington, DC 20370

Head,
Military Compensation
Policy Branch, NAVOP 134
Washington, DC 20370

Head,
Workforce Information Section,
NAVOP 140F
Washington, DC 20370

Head,
Family Support Program
Branch, NAVOP 156
1300 Wilson Blvd. Room 828
Arlington, VA 22209

Head,
Economic Analysis Branch
NAVOP 162
Washington, DC 20370

Head,
Manpower, Personnel, Training &
Reserve Team, NAVOP 914D
5A578, The Pentagon
Washington, DC 20350

Assistant for Personnel Logistics
Planning
NAVOP 987H
5D772, The Pentagon
Washington, DC 20350

Leadership Management Education and
Training Project Officer
Naval Medical Command
Code 05C
Washington, DC 20372

Technical Director
Navy Health Research Center
P.O. Box 85122
San Diego, CA 92138

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

Psychologist
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, N.W.
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. James W. Pellegrino
University of California,
Santa Barbara
Department of Psychology
Santa Barbara, CA 93106

Military Assistant for Training
and Personnel Technology
OUSD (R&E)
Room 3D129
The Pentagon
Washington, DC 20301

Administrative Sciences Dept.
Naval Postgraduate School
Monterey, CA 93940

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93940

Department of Computer Sciences
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Bernard Rimland
Navy Personnel R&D Center
San Diego, CA 92152

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

Dr. Robert Sasmor
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

END

DTIC

9 — 86